DOCUMENT RESUME

ED 272 841 CS 008 486

AUTHOR Rankin, Earl F.; Helm, Phoebe
TITLE The Validity of Cloze Tests in Relation to a Psycholinguistic Conceptualization of Reading Comprehension.
PUB DATE 86
NOTE 16p.; "Forum for Reading" is the journal of the College Reading Improvement Special Interest Group of the International Reading Association.
PUB TYPE Viewpoints (120) -- Journal Articles (080)
JOURNAL CIT Forum for Reading; v17 n2 p46-59 Spr-Sum 1986

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cloze Procedure; Elementary Secondary Education; Higher Education; *Measurement Techniques; *Reading Comprehension; *Reading Research; *Reading Tests; *Test Validity

ABSTRACT

Reading comprehension, viewed from a psycholinguistic perspective, is a process of mapping new information into existing knowledge structures and is similar to general thinking, reasoning, and problem solving. Reading comprehension is an active process--a dialogue between writer and reader--and an integration of the printed page and the reader's mind. Despite publication of much research supporting the validity of cloze tests as measures of reading comprehension over a period of about 30 years, both the content and the construct validity of cloze tests are still questioned. Further research and a review of the literature indicate that the content validity of cloze tests as measures of a meaningful construct of comprehension is stronger than for conventional reading comprehension tests, which have a higher validity only because of the uncritical acceptance of conventional reading test formats as models for the valid measurement of reading comprehension. Furthermore, the construct of reading comprehension measured by cloze tests is more theoretically and empirically meaningful than notions of reading comprehension embodied in conventional reading tests. A three-page list of references concludes the document. (SRT)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED272841

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

The Validity of Cloze Tests in Relation to a Psycholinguistic Conceptualization of Reading Comprehension

Earl F. Rankin

Phoebe Helm

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Rona F. Flippo

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS008486

The validity of cloze tests in relation to a psycholinguistic conceptualization of reading comprehension

Earl F. Rankin
University of Kentucky
Lexington, Kentucky

Phoebe Helm
Triton College
River Grove, Illinois

Part I — Content Validity

The ease and rapidity with which cloze tests and exercises can be constructed by both college level reading specialists in developmental reading and study skills programs and classroom instructors in subject matter areas should make them widely used in colleges and universities. They can serve a variety of useful purposes such as determining the relative difficulty of texts and other materials, measuring students' specific comprehension of textbook assignments and/or reading improvement materials selected from specified areas of student interest from magazines and newspapers, and estimating background knowledge of individual readers on topics to be used for training purposes, etc.

Although the potential exists for many practical uses of cloze materials, they are not in fact used by a good many college personnel. This may be in part because they do not appear to be measures of reading comprehension similar to other conventional measurement procedures that have long been used and accepted both in higher educational settings and in elementary and secondary classrooms.

Despite publication of many research papers supporting the validity of cloze tests as measures of reading comprehension over a period of about thirty years, the question is still asked, "Does a cloze test really measure reading comprehension?" This question is asked by many people who have read neither the positive critiques on this issue in the literature by writers such as Taylor (1957), Rankin (1978), Bormuth (1966), Oller (1979), Rankin (1983), nor the negative critiques by writers such as MacGinitie (1960), Carroll (1972), Tuinman, Blanton, and Gray (1975), Alderson (1978), Shanahan, Kamil, and Tobin (1982), and Shanahan and Kamil (1982). Why does this question remain despite the overwhelming amount of empirical evidence in the literature over many years supporting the validity of cloze tests and more than a thousand studies utilizing cloze tests in research on many aspects of language?

The answer to this question may lie (a) in the lack of "face validity" of cloze tests when contrasted with conventional reading tests and (b) the superficial reliance upon "criterion-related validity" studies unrelated to any general theory of the construct "reading comprehension." Conventional reading tests have high face validity only because people have become accustomed to that type of test format, and their validity relies primarily upon correlations with various validity criteria whose construct validity has not been determined.

"Face validity" is a term which refers to the acceptability of a test by the users of that test. The lack of "acceptance" of a test does not really represent limited validity unless the negative evaluation of the test is so extreme that it influences the test taking behaviors of individuals taking a test. In any case, it is obvious from visual inspections of cloze tests and by introspective and retrospective reports of individuals taking cloze tests that they are clearly not measuring the same thing as conventional reading tests despite

any apparent statistical evidence to the contrary. Typical questions on conventional reading tests measure such skills as recognizing main ideas, recalling details, and applying many types of literary interpretation. If one assumes that conventional reading tests do measure the construct "reading comprehension", then from the standpoint of face validity one might easily jump to the conclusion that whatever cloze tests measure, they certainly must not be measuring reading comprehension.

The large amount of research on cloze test criterion-related validity demonstrating high positive correlations between cloze test results and other reading test data indicates that cloze tests and conventional reading tests are measuring something in common, but the nature of that "something" is unclear. After all, correlations exist between height and weight, but these correlations do not prove that a yardstick and a bathroom scale measure the same thing. Therefore, since the empirically established validity of both cloze tests and conventional reading tests is based largely upon correlational studies, the question remains about what either type of test is in fact measuring.

Paradoxically, cloze tests have low face validity and high criterion-related validity, and this fact prevents us from understanding more fully just what trait cloze tests are measuring. The resolution of this paradox lies, we believe, in a critical analysis of the construct "reading comprehension" in relation to psycholinguistic views of reading and a careful consideration of two other types of test validity, i.e., "content validity" and "construct validity." Due to the complexity of this problem, the content validity of cloze tests will be examined in the first major section of this paper; the construct validity of cloze tests will be considered in the second section. In the light of these studies, based upon a combination of logical analysis (for content validity) and empirical evidence (for construct validity), it will be shown that (a) cloze tests measure a variable which is a necessary but not a sufficient prerequisite for adequate performance by readers on conventional reading tests and (b) the construct of reading comprehension measured by cloze tests is more theoretically and empirically meaningful than notions of reading comprehension embodied in conventional reading tests.

A Psycholinguistic Conceptualization of Reading Comprehension

Reading comprehension is a complex process and has been viewed by scholars in many ways. Huey (1908) believed that to understand reading would be to understand the mysteries of the human mind. Since Huey's early work, only a little progress has been made in understanding these mysteries. However, more recent developments in psychology and linguistics seem to hold some promise for building a better understanding of this mysterious process. The writings of Smith (1975, 1978), Pearson and Johnson (1978), and Oller (1979) embody significant psycholinguistic views of reading. Each of these authorities has formulated somewhat different views of reading which integrate basic concepts in cognitive psychology and linguistics, and yet there is a common core of basic assumptions which runs throughout the thoughts of each writer. Unfortunately, space does not permit a consideration of the differences among them. However, a careful study of their views leads to a set of common generalizations about the nature of reading comprehension which will be useful in studying the content and construct validity of the cloze procedure. These generalizations are as follows:

1. Reading comprehension involves the processing of information from printed (or written) sources.
2. Reading comprehension is similar to general thinking, reasoning and problem solving.

3. Reading comprehension is an active process—a dialogue between a writer and a reader, in which effective communication depends upon expectancies of both the sender and the receiver.
4. Reading comprehension is dependent upon the integration of two contexts or information sources, that is, (a) information on the printed page (linguistic) and (b) information inside the head of the reader (extra-linguistic or pragmatic).
5. The processing of information through reading depends greatly upon the ability to make inferences.
6. Reading comprehension is a process of pragmatically mapping new information into existing structures through assimilation or accommodation.

The following definition of the construct "reading comprehension" is adopted from Oller's view of language and is viewed by the writers as being compatible with the psycholinguistic views of all three writers:

Reading comprehension is an interactive communication process between a writer and a reader in which the reader uses a "pragmatic expectancy system"[1] (or grammar) to process information in the message by utilizing normal contextual constraints in the language and relating these linguistic sequences to extra-linguistic events.

Logical Evidence for Content Validity

Content validity of a test is determined through a process of rational judgment which involves matching the content of a test with the skills or knowledge which presumably constitute the "content" or "domain" measured by that test. The content validity of cloze tests will be explored through logical analysis of sample cloze test items in relation to four basic assumptions which are consistent with the previously defined psycholinguistic construct of reading comprehension. These four assumptions incorporate concepts from all six of the previously listed generalizations. They represent views of reading which overlap to some extent and are not intended to represent mutually exclusive categories.

First Assumption. Reading comprehension as a process involves the extraction of information from a printed or written source prior to the utilization of this information.

This assumption points to the difference between reading comprehension viewed as a process and as a product. As a holistic process, reading comprehension can be thought of as what readers do as *they read.* Whereas reading comprehension as a product might be defined as what readers do *after they read* as a result of having processed information in the text. For example, let us consider the following statement, "Please, pass the salt." If, after having been asked the meaning of the statement, the reader responded by saying, "You asked me to hand you the salt," one would have an indication that the information had indeed been processed. However, it is possible for one to have processed the information involved in the statement and not have the information or skill necessary to respond to other types of questions frequently found in reading tests such as "What is the definition of the word 'pass'?" or "Why were you asked to pass to salt?". In these latter types of questions, comprehension as a process and a product are confounded as in many conventional tests.

Smith (1978) views reading comprehension as a process of making predictions and utilizing redundancy to confirm or deny those predictions until the reader's uncertainty reaches zero. It can be shown that cloze items measure these intellectual functions. Why do cloze-type items measure reading comprehension as a process? What exactly is taking place when a

reader responds to an item on a cloze test? Consider the following example:

> We like to keep our houses warm in the _____ and, in contrast, to keep them _____ in the summer.

On the basis of the incomplete information in the example above, the reader must project a word for each blank which will complete the sequence. To do so, the reader utilizes not only the information in the statement, but also his long-term memory regarding comfortable temperatures for homes, to infer that most persons would like their homes to be warm in the winter and cool in the summer. The reader may also use his knowledge of language or structure in writing to determine that the first blank would be filled in with a word which is parallel to "SUMMER" ("winter" in this case) and similarly, the second blank should be filled in with a word which is parallel to "WARM" ("cool"). In completing cloze items on a pre-cloze test, (a situation in which the reader has not read the unmutilated passage), as in the example above, the reader is not using long-term memory for the words in the passage, neither is he/she guessing wildly. Cloze tests intermittently sample the inferences (syntactic and semantic) the reader is making during the process of reading the passage. Thus, cloze scores become an indicator of readers' ability to process the information in the text, not the ability to respond to specific questions which may not have been of any concern to them at the time when they were reading the text (Rankin & Thomas, 1980). Weaver (1965) apparently shared this view a good many years ago when he stated that the cloze procedure, "... enables us to get a close-up view of what is occurring at particular points in a language passage." (p. 17)

Second Assumption. Reading comprehension is an active process—a dialogue between a writer and a reader in which the effectiveness of the communication is dependent upon the expectancies of both the sender and the receiver.

Reading is not a passive activity; the reader must make a substantial and active contribution if he/she is to make sense of print. While the reader is obviously constrained by what the writer is saying, writers are also constrained by what they wish to say and how they wish to say it. In addition, the writer is constrained by the intended audience, by the amount of information he/she assumes them to already have. The writer is also constrained by the reaction or desire he/she wishes to create in the reader.

Furthermore, for communication to be effective, the language patterns of the writer and the reader must be compatible. Just as writers plan what they are going to say and how they are going to say it, the reader anticipates much of what the writer is likely to say. Oller (1979) explained the role of expectancy in relation to oral language in a way which is equally applicable to reading:

> Whenever we say anything at all we leave a great deal more unsaid. We depend largely for the effect of our communications not only on what we say but also on the creative ability of our listeners to fill in what we have left unsaid. The fact is that a normal listener supplies a great deal of information by creative inference, and in a very important sense is always anticipating what the speaker will say next. Similarly, the speaker is always anticipating what the listener will infer and is correcting his output on the basis of feedback received from the listener. Of course, some language users are more skillful in such things than others. (p. 20)

An example of how a reader anticipates what a writer is likely to say can be seen in the following cloze items used by Taylor (1953): "I heard a _____ bark." The reader would be likely to fill in the word "dog." However, if the statement were, "For the first time, I heard a _____ bark," the reader would be less likely to project the word "dog." He/she might

think, "Who is this person who has never heard a dog bark?" However, if the context of this statement included the discussion of a voyage to the Pribilif Islands, the reader would be likely to produce the word: "seal" (p. 418). The above statements illustrate the influence of the writer's expectancies of the reader and how readers may revise their expectancies as a result of the information provided by the writer. Cloze tests repeatedly sample the likenesses of the pragmatic expectancy grammar of the reader and the writer and, thus, sample the effectiveness of communication.

Third Assumption. Drawing inferences is essential to the processing of information through reading.

Basic to understanding this assumption and analyzing cloze tests as measuring this process is an understanding of the role of inference in information processing. When information is processed, uncertainty is reduced and perceptual and cognitive gaps which existed prior to the processing of the information are completed. Smith (1975, 1978) viewed reading as one form of information processing and defined comprehension as a state in which uncertainty has reached zero.

The reader is able to reduce uncertainty through the use of redundancy (information from more than one source). Two primary sources of redundancy in reading are (1) information inside the head and (2) information on the page. The reader utilizes these two sources of information to make inferences (predictions) regarding what ideas are likely to occur next in a given passage. The reader then searches the text to confirm, deny or revise these inferences. This process of drawing inferences is continued throughout the reading act; thus, completing the perceptual and cognitive gaps which existed prior to reading the passage.

In taking a cloze test the reader responds to information from several sources (i.e., redundancy) by making inferences among words to be considered for filling in the gaps within the message. For example, consider hypothetically the process by which cloze responses might be made in completing the following statement:

The cautious deer walked _____ the clearing, then looked _____ in each direction before _____ from the pond.

This item and the following explanation is adapted from Pearson and Johnson (1978, p. 173). To complete the cloze gaps in this statement, the readers make many inferences. First, they have inferred that the deer is in the woods, high grass or some other relatively secluded area. Secondly, they have inferred that the pond is in the clearing and that the deer wants to drink from the pond. Thirdly, they have inferred that there is some measure of danger to be considered when a deer enters a clearing to drink from a pond. Readers are able to make these inferences because of their previous experience or knowledge regarding the characteristics and behaviors of deer which has been stored in their long term memory. Readers expect the message in the print to be compatible with what they know about deer. By utilizing this redundancy together with the redundancy in the passage, the readers are able to reduce much of the uncertainly in cloze gaps.

This redundancy in print is categorized as sequential or distributional. Sequential redundancy includes the syntactic and semantic constraints imposed by the language. That is, certain types of words follow other words, hold certain positions, and fulfill certain functions within a sentence. Distributional redundancy refers to the fact that some words occur more frequently in the language, regardless of context, than do other words. Examples of high frequency words include "a, and, the, but, and is." Skilled readers have both sequential and distributional redundancy rules programmed in their long term memory.

In looking at the first cloze gap in the above example, the reader may use sequential redundancy to limit his considerations to function words. By doing this, he/she has greatly reduced the uncertainly surrounding this gap because the number of function words in the English language is relatively small. The reader could, at this point, use his/her knowledge of distributional redundancy to limit his/her consideration to high frequency function words. For example, the words "in, into, around and beyond" are statistically more likely to occur than the word "throughout." By using the semantic constraints in the passage, as well as one's previous knowledge of deer, the reader further reduces uncertainly and thus restricts inferences to even fewer function words. For example, the function words "around, beyond and into" would satisfy the syntactical constraints; however, the reader has already inferred that the pond is in the clearing and that the deer wants to drink from it. Therefore, he/she would not be as likely to choose the words "around or beyond" because they would not get the deer to the pond. Thus, the reader would choose the word "into" to complete the first cloze gap. By using both types of redundancy, the reader continuously makes inferences, reducing uncertainty bit by bit until the gap is filled. This process might be continued in a similar fashion to complete the second and third cloze gaps in the sentence with the words "carefully" and "drinking" respectively.

It should be noted that while the processing of information through reading is very much dependent upon the process of inference, using multiple and overlapping sources of information, the able reader is seldom conscious of the process. The cloze procedure artifically imposes gaps for the reader to fill in by mechanically deleting words from the message and continously sampling the accuracy of the reader's inferences throughout a passage. Thus, it provides some insights into the data sources which must have been used by the reader to complete the cloze gaps.

Fourth Assumption. Reading comprehension is a process in which the reader makes textually explicit, textually implicit, and scriptally implicit inferences as he reads.

It should be pointed out that cloze test items do not directly measure textually explicit inferences as defined by Pearson and Johnson (1978); the answers to cloze items are never obviously and clearly stated in the text. However, cloze tests are sensitive to textually and scriptally implicit inferences.

The following is an example of a cloze item which would be sensitive to textually implicit inferences: "Tom struck it rich in Texas; _____ discovered oil in his backyard." While the completion of the sequence with the word "he" seems quite simple, it does require that the reader recognize that Tom is a male, and that Tom discovered oil. Also, since the author used the proper noun "Tom" in the preceding clause, he/she is more likely to have used the pronoun "he" in the subsequent clause. Thus, cloze items can measure rather subtle aspects of drawing inferences.

An example of a cloze item which required the reader to utilize his script (information stored in the long term memory regarding longer units of discourse or events) can be found in the following statement by Weaver (1965), "The young boys went downtown their first day in the new town. They were chased home by a gang of roughnecks—rock-throwing roughnecks. The next day they tried again, the same thing happened, and again the next day. They _____ going to town" (p. 7). While it is very easy within the context of this passage to supply the word "stopped," the response is not found in the passage. Supplying the word "stopped" was easy because it came from the "scripts" in the reader's head regarding roughnecks, especially rock-throwing roughnecks. The reader probably also called on his/her

scripts concerning the behavior which might be expected of boys who had been chased home by rock-throwing roughnecks two or three days in a row.

Although cloze items might look like "fill in the blank" items calling only upon rote memory, they most often measure inferential processes. It seems clear from the examples above that although cloze is not sensitive to textually explicit inferences, it is sensitive to the higher levels of inference as categorized by Pearson and Johnson (1978).

Summary and Conclusions:
Part I — Content Validity

The authors have attempted to prove, through a rather detailed logical analysis of sample cloze test items in relation to a psycholinguistic construct of reading comprehension, that cloze tests do, indeed, measure reading comprehension as a process. Taylor (1954) derived the strategy for constructing cloze tests from information theory and language theory. Conventional reading tests are not closely related to psycholinguistic theories of reading. Their questions conform to pedagogical conventions and traditional assumptions about skills presumably involved in reading comprehension. As has been pointed out, conventional reading tests measure some of the products or uses of information that has been extracted by the reader from the printed page, whereas cloze tests measure the ability to extract the information more directly. The former is a necessary but not a sufficient condition for the latter to occur. In neither case, of course, do cloze or conventional reading tests measure directly an underlying trait called "reading." Responses to any type of test item consist of behaviors which are believed to reveal the amount of the underlying trait which the test taker is presumed to possess. In this sense, both cloze and conventional test item responses are several steps removed from the phenomenon being measured. However, in the opinion of the writers, cloze test items elicit responses which can be shown through a theoretical analysis to bear a closer relationship to the underlying trait of information processing involved in all reading than is possible with objective test questions. Therefore, the apparent lack of cloze test face validity is really due to the uncritical acceptance of conventional reading test formats as models for the valid measurement of reading comprehension. Although both types of tests have acceptable criterion-related validity, the evidence for the content validity of cloze tests as measures of a meaningful construct of comprehension is stronger than for conventional reading comprehension tests.

Part II — Construct Validity

In Part I, the "content validity" of cloze test items was examined through a process of logical analysis in relation to several basic assumptions underlying a psycholinguistic construct of reading comprehension. In Part II, the "construct validity" of cloze tests is evaluated by citing empirical evidence in support of hypotheses drawn from the same psycholinguistic construct of reading. The goal of both strategies for evaluating validity was the same, that is, evaluating the ability of cloze tests to measure a construct of reading comprehension, that is the processing of information through reading. However, unlike the "armchair" logical analysis of content validity, the empirical assessment of construct validity must meet rigorous criteria of scientific research in testing hypotheses stemming from the therory or construct underlying the test.

The search for construct validity of a test involves asking questions such as the following: (a) What do scores on a cloze test mean?, or (b) Do cloze test results correspond to an underlying trait of reading comprehension? In seeking answers to these questions, empirical evidence was sought from studies in the literature concerning the extent to which cloze test

results relate to other variables in a manner predicted from the nature of the construct being measured. Several hypotheses were derived from the psycholinguistic construct of reading comprehension discussed in the preceding section and tested using the results from previously published works. These hypotheses were stated in an "If . . . then" paradigm. For example, if cloze tests measure "x," then scores on cloze tests will be related to "y" in a predictable manner.

Hypotheses

The following hypotheses were formulated by the writers for the purpose of studying the construct validity of cloze tests as a measure of reading comprehension:

H_1 If reading comprehension is processing information from printed sources, then cloze scores should be a reflection of the amount of information processed as indicated by the following evidence:

a. Deletions of high frequency words (low information) in cloze tests should be more easily replaced than low frequency words (high information).

b. Cloze scores should be higher for easy passages than for difficult passages.

c. Cloze scores should be higher for post-cloze tests than for pre-cloze tests.

d. Cloze scores should be higher in passages where more context is supplied (less frequent deletions) than in passages which contain less context (more frequent deletions).

H_2 If the process of reading comprehension is similar to the process involved in general thinking, reasoning and problem solving, then, cloze tests should correlate highly with tests of intelligence, reasoning, and problem solving.

The following discussion focuses on the empirical findings of selected cloze studies as they relate to these hypotheses.

Hypothesis Ia. High frequency words (low information) should be more easily replaced in cloze tests than low frequency words (high information).

High frequency words tend to be function words and are referred to by LaBerge and Samuels (1976) as being automatically coded or processed. In his very early studies of the cloze procedure, Taylor (1957) found high frequency words (pronouns, articles, conjunctions and verb auxiliaries) to be more easily replaced by readers than were content words. Fillenbaum et al, (1963) found subjects correctly replaced 65% of the function words in a cloze test, but replaced only 32% of the content words correctly.

In a reanalysis of Bormuth's (1966) data, Finn (1977-1978) found the percentage of correct cloze responses (cloze easiness) to correlate .55 with the standard frequency index of the English language. Finn concluded that frequently used words are processed easily and rapidly because they are low in information; that is, if the reader is 99% certain of what the next word will be, the actual occurrence of the word yields virtually no information (p. 513). High frequency words are replaced more easily in cloze tests because they are fewer in number (distributional redundancy) and tend to be highly constrained both syntactically and semantically (sequential redundancy); therefore, fewer alternatives are available to the reader and less uncertainty remains to be reduced by the reader at such cloze gaps.

Hypothesis Ib. Cloze scores should be higher for easy passages than for difficult passages.

This hypothesis is based on the assumption that easier passages are more redundant than more difficult passages and that they contain a smaller percentage of content words (and a higher percentage of function words) than do the more difficult passages. The basic assumption underlying

Taylor's use of the cloze procedure to study the readability of a passage or text was that the higher the cloze score, the more readable (easier) the text was considered to be. Taylor (1953) found that cloze test results produced rankings of passage difficulty similar to the Flesch readability formula and the Dale-Chall formula. Taylor's findings that subjects scored higher on easier passages have been replicated by many researchers in studying the reading difficulty of tests.

Finn (1977-1978) hypothesized that any word deletion would be easier to replace in a passage of low difficulty than in a passage of high difficulty. He found that cloze easiness correlated (-.38) with passage difficulty. Through further analysis of this data, Finn found these correlations to range from (-.54) for rare words to (-.34) for easy words. Greater passage redundancy, of course, makes for more accurate predictions for both types of words.

Hypothesis lc. Cloze scores should be higher for post-cloze tests than for pre-cloze tests.

Taylor (1957) and Rankin (1959) found gains between pre- and post-reading cloze tests significant at the $p<.001$ level. Greene (1964), however, did not find significant differences in the pre-post reading cloze test scores of his subjects. This finding is atypical for such comparisons. It may have been due to the use of a different item selection procedure which limited cloze items to those which were predictable for the available content rather than using an every nth deletion procedure.

In studying the use of cloze tests to measure information gained in science courses, Charles (1978) found subjects scored significantly higher on cloze tests constructed from the materials previously used in class; i.e., the experimental group scored higher on the individualized materials while the control group scored higher on passages taken from the textbook.

Rankin and Thomas (1980) found that exposure to the unmutilated text increased cloze scores, but the effect was greater for content words than for function words. This finding was seen by the authors as consistent with psycholinguistic views of reading since function words are high frequency, low information words, they would not be expected to be influenced by familiarity (having previously read the unmutilated passage) as much as content words which are high information, low frequency words.

Hypothesis ld. Cloze scores should be higher in passages where more context is supplied (less frequent deletions) than in passages of less context (more frequent deletions).

The question of sensitivity of cloze scores to contextual constraints beyond a few words on either side of the cloze gap has been the subject of many investigations and much controversy. Aborn, Rubenstein, and Sterling (1959) found that increasing the context beyond ten words did not increase cloze scores. Similarly, MacGinitie (1960) found that increasing the context beyond five words did not increase cloze test scores. Tuinman, Blanton, and Gray (1975) concluded that cloze scores were influenced only by local redundancy. More recently, Shanahan, Kamil, and Tobin (1982), and Shanahan, and Kamil (1982) have carried out studies supporting the hypothesis that cloze items are not sensitive to clues beyond a few spaces on either side of the cloze gap. (It should be noted that Henk (1982) has pointed out several deficiencies in the Shanahan, Kamil, and Tobin paper.)

Conversely, Marshall (1970) found subjects could take advantage of increasing amounts of context in completing cloze items embedded in fragments, isolated sentences and connected discourse. Ramanauskas (1972) found that scrambling the sentence order within a passage depressed cloze test scores and concluded that subjects could take advantage of context beyond the sentence in responding to cloze items. Weaver (1965) found that

when increased context was provided, subjects used greater amounts of context in completing cloze gaps and also noted that their rate of production of words was much faster with the increased amount of context than without it. In a study using aural cloze tests, Nutter (1974) found the every 10th deletion pattern to be easier for his subjects than was the every 5th deletion pattern. Finn (1977-1978) found that low frequency words, which appeared frequently in a passage (high text frequency), were processed similar to high frequency words as the reader progressed through the passage. For example, the word "cotton" appeared in the first, sixteenth and twenty-eighth position in the passage. The word "cotton" was correctly replaced in the first position by 10% of the subjects; in the sixteenth position by 22% and in the twenty-eighth position by 50% of the subjects. These findings are not consistent with the results of previously cited studies indicating that cloze items are only sensitive to local contextual constraints.

Different methodologies employed in various studies such as use of different word deletion ratios, use of easy vs. difficult materials, and use of natural vs. artificial language in cloze test construction may explain inconsistencies in the findings of different publications.

Using cloze tests based upon reading materials in keeping with the reading ability of the subjects, Rankin and Thomas (1980) found increased amounts of context to facilitate subjects' ability to correctly restore content words, but to depress the restoration of function words. This important finding was suggested by the authors as possibly explaining the failure of some studies to find effects of greater context upon word production. Since contextual constraint studies have tended to use easy materials (which presumably contained a larger portion of function words than would be found in more difficult materials), Rankin and Thomas suggest that the results of such studies would therefore not be expected to reveal the effects of greater context upon responses to cloze items. Their findings and conclusions are in keeping with the theoretical analysis of the role of context in making global vs. focal predictions in the psycholinguistic views of Smith (1978). This distinction has not previously been considered in other contextual constraint studies.

Although this issue is by no means "clozed," the writers believe that the evidence in several studies supporting the hypothesis that greater context tends to increase cloze scores is convincing and, for previously stated reasons, this evidence has not been disproved by conflicting results of other investigations.

Hypothesis 2. If the process of reading comprehension is similar to the processes involved in general thinking, reasoning and problem solving, then, cloze tests should correlate highly with tests of intelligence, reasoning and problem solving.

Both Taylor (1956) and Rankin (1957) found pre- and post-cloze tests to be significantly related to intelligence. Weaver, White and Kingston (1968) reported studies in which pre-cloze tests correlated relatively high in the .70's with standard tests of intelligence. Hafner (1966) reported that he found poor readers, as measured by cloze tests, to be less intelligent and less able to reason well, than better readers. In a factor analytical study of cloze tests, Kohler (1966) found logical reasoning, wide range vocabulary and inference (described as less formal reasoning) to be consistently and significantly related to cloze tests. In studying cloze and dictation tasks as predictors of intelligence and achievement scores, Stump (1977) found that scores on the Lorge-Thorndike Intelligence Test, the Iowa Test of Basic Skills, a cloze test and a dictation test to be highly correlated. Oller and Perkins (1978) reported data which seemed to indicate that a single factor

(language proficiency) accounted for virtually all of the reliable variance in I.Q. and achievement tests. Oller (1979) concluded that, "the g factor of intelligence is indistinguishable from global language proficiency" (p. 456). Whether or not this is the case, it should be noted that the correlations cited in this section provide, at best, weak and indirect evidence for construct validity of cloze tests per se. Virtually all verbal tests correlate positively with intelligence and verbal problem solving. On the other hand, if cloze tests results failed to display such predictable relationships, their validity would obviously be called into question.

Summary and Conclusions:
Part II — Construct Validity

This study of construct validity has cited a variety of empirical research findings which are consistent with several hypotheses drawn from a psycholinguistic construct of reading comprehension from which cloze tests are assumed to be derived. In the writers' judgment, the converging evidence from different sources clearly supports cloze test construct validity. It should be noted however that these studies were not themselves investigations of construct validity. Rather, this paper brings together findings from diverse sources which were carried out at different times to investigate various problems related to cloze test performance.

It would be preferable to carry out a series of coordinated construct validity studies testing a series of hypotheses logically derived from psycholinguistic theory on the same or randomly equated groups of subjects. Nevertheless, it is believed that the converging evidence from different published sources cited in the present study both (a) adds strength to the contention that cloze tests are measures of a psycholinguistic construct of reading comprehension and (b) will stimulate future construct validity studies on both cloze tests and conventional reading tests. Strong evidence for the construct validity of most reading tests is not often found. It is only as more careful study is made of the construct validity of all reading tests that we can have greater confidence in our understanding of the true nature of the phenomenon that we are presumably measuring.

Footnote

[1]Oller (1979) defines "pragmatic expectancy system" as, ". . . a psychologically real system that sequentially orders linguistic elements in time and in relation to extra-linguistic contexts in meaningful ways." (p. 34)

References

Alderson, J. C. (1978). Review of cloze procedure (Ebbinghaus sentence completion method) as applied to reading. In O. K. Buros (Ed.). *The Eighth Mental Measurements Yearbooks*, (pp. 1171-1174). Highland Park: NJ.

Aborn, J., Rubenstein, H., & Sterling, T. D. (1959). Sources of contextual constraints upon words in sentences. *Journal of Psychology, 57*(3), 171-180.

Bormuth, J. R. (1962). Cloze tests and measures of readability. Unpublished doctoral dissertation, Indiana University, Bloomington.

Bormuth, J. R. (1966). Readability: A new approach. *Reading Research Quarterly, 1*(3), 79-132.

Carroll, J. B. (1972). Defining language comprehension: Some speculations. In J. B. Carroll and R. O. Freedle (Eds.). *Language comprehension and the acquisition of knowledge*. Washington: V. H. Winston and Sons.

Charles, E. (1978). An investigation of the use of cloze tests to compare gain scores of students in science who have used individualized science materials and those who have used traditional textbook materials. Unpublished doctoral dissertation, Lehigh University, Bethlehem, PA.

Fillenbaum, S., Jones, L., & Rapoport, A. (1963). The predictability of words and their grammatical classes as a function of rate of deletion from a speech transcript. *Journal of Verbal Learning and Verbal Behavior, 2*(2), 186-194.

Finn, P. (1977-1978). Word frequency, information theory, and cloze performance: A transfer feature theory of processing in reading. *Reading Research Quarterly, 13*(4), 508-537.

Greene, F. (1964). A modified cloze procedure for assessing adult reading comprehension. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Hafner, L. (1966). *Journal of Reading, 9*(6), 415-421.

Henk, W. A. (1982). Commentary: A response to Shanahan, Kamil, and Tobin: The Case is not yet clozed. *Reading Research Quarterly, 17*(4), 591-595.

Huey, E. (1968). *The psychology and pedagogy of reading*. New York: McMillen, 1908 and Cambridge, Massachusetts: MIT Press.

Kohler, E. (1966). An investigation of cloze scores in terms of selected cognitive variables. Unpublished doctoral dissertation, Florida State University, Tallahassee.

LaBerge, D., & Samuels, S. (1976). Toward a theory of automatic information processing in reading. In H. Singer and R. Ruddell (Eds.). *Theoretical Models and Processes of Reading* (2nd ed.). International Reading Association, 548-579.

MacGinitie, W. (1960). Contextual constraints in English prose. Unpublished doctoral dissertation, Columbia University, New York.

Marshall, W. (1970). Contextual constraint on deaf and hearing children. *American Annals of the Deaf, 115*, 682-689.

Nutter, B. (1974). Presentation methods, deletion patterns and passage types for use with aural cloze. Unpublished doctoral dissertation, University of Arizona, Tucson.

Oller, J. (1979). *Language Tests at School*. London: Longman.

Oller, J., & Perkins, K. (Eds.). (1978). *Language in Education: Testing the Tests*. Rowley, MA: Newberry House.

Pearson, D., & Johnson, D. (1978). *Teaching Reading Comprehension*. Chicago: Holt, Rinehart and Winston.

Ramanauskas, S. (1972). Contextual constraints beyond a sentence on cloze responses of mentally retarded children. *American Journal of Mental Deficiency, 77*(3), 338-345.

Rankin, E. (1957). An evaluation of the cloze procedure as a technique for measuring reading comprehension. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Rankin, E. (1959). The cloze procedure—its validity and utility. In O. S. Causey and W. Eller (Eds.), *Starting and improving college reading programs: Eighth Yearbook of the National Reading Conference*, (pp. 131-144). Fort Worth, TX: National Reading Conference.

Rankin, E. (1978). Characteristics of the cloze procedure as a research tool in the study of language. In P. D. Pearson and J. Hansen (Eds.), *Reading: disciplined inquiry in process and practice:* Twenty-seventh Yearbook of the National Reading Conference, (pp. 148-153). Clemson, SC: National Reading Conference.

Rankin, E., & Thomas, D. (1980). Contextual constraints and the construct validity of the cloze procedure. In M. L. Kamil and A. Moe (Eds.), *Perspectives on reading research and instruction:* Twenty-ninth Yearbook of the National Reading Conference, (pp. 47-55). Washington, DC: National Reading Conference.

Rankin, E. (1983). An evaluation of the ecological item format and ecological validity of the degrees of reading power test. *Eastern Kentucky University Educational Review*, 7(1), 8-18.

Shanahan, T., & Kamil, M. L. (1982). The sensitivity of cloze to passage organization. In J. A. Niles and L. A. Harris (Eds.), *New inquiries in reading research and instruction*. Thirty First Yearbook of the National Reading Conference (pp. 204-208). Rochester, NY: National Reading Conference.

Shanahan, T., Kamil, M. L., & Tobin, A. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly*, *17*(2), 229-255.

Smith, F. (1975). *Comprehension and learning*. New York: Holt, Rinehart and Winston.

Smith, F. (1978). *Understanding reading*. (2nd ed.). New York: Holt, Rinehart and Winston.

Stump, T. (1977). Cloze and dictation tasks as predictors of intelligence and achievement scores. Paper presented at the Eleventh Annual Meeting of TESOL, Miami, Florida, 1977. In Oller and Perkins (Eds.) (pp. 36-63). *Language in education: Testing the tests*. Rowley: MA, Newberry House.

Taylor, W. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, *30*(4), 415-433.

Taylor, W. (1954). Application of cloze and entrophy measures to the study of contextual constraint in samples of continuous prose. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Taylor, W. (1956). Recent developments in the use of cloze procedure. *Journalism Quarterly*, 33(1), 42-48.

Taylor, W. (1957). Cloze readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology*, *41*(1), 19-26.

Tuinman, J., Blanton, W., & Gray, G. (1975). A note on cloze as a measure of comprehension. *Journal of Psychology*, *90*(2), 159-162.

Weaver, W. (1965). Theoretical aspects of the cloze procedure. In E. Thurston and L. Hafner (Eds.), *The philosophical and sociological bases of reading:* Fourteenth Yearbook of the National Reading Conference. (pp. 115-132.). Fort Worth, TX: National Reading Conference.

Weaver, W., White, W., & Kingston, A. (1968). Affective correlates of reading comprehension. *Journal of Psychology, 68* (1), 87-95.